Camera-based PHM method in rotating machinery equipment micro-action scenarios

Indexed by:
Web of Science Group

Junfeng An[a,b*], Jiqiang liu[a], Hao Zhen[a], Mengmeng Lu[c]

[a] Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing, 100044, China
[b] Jinan Rail Transit Group Co., LTD, Jinan, 250000, China
[c] Shandong Labor Vocational and Technical College, Jinan, 250103, China

Highlights

- A new PHM method which combines image processing and deep learning.

- Condition monitoring, anomaly detection, defect early warning using micro-action.

- Rmcad Framework, TSML-net, 1DCNN.

- Visualization of categories without dimension reduction.

- Using a camera as a sensor, sensing the data of opening degree, anomaly label, and health label.

Abstract

The health operation of rotating machinery guarantees safety of the project. To ensure a good operating environment, current subway equipment inspections frequency is high, resulting in a waste of resources. Small abnormal changes in mechanical equipment will also contribute to the development of mechanical component defects, which will ultimately lead to the failure of the equipment. Therefore, mechanical equipment defects should be detected and diagnosed as soon as possible. Through the use of graphic processing and deep learning, this paper proposes Rmcad Framework with three aspects: condition monitoring, anomaly detection, defect early warning. Using a network algorithm, this paper proposes an improved model that has the characteristics of two-stream and multi-loss functions, which improves the accuracy of detection. Additionally, a defect warning method is constructed to improve the perception ability of equipment before failure occurs and reduce the frequency of frequent maintenance by detecting anomalies according to the degree of opening.

Keywords

deep learning, condition monitoring, Rmcad, anomaly detection, defect early warning

1. Introduction

The use of rotating machinery has been widespread in aviation, rail transport, and petrochemical engineering. As a result of wear, speed, load, external shock, and other effects in operation, performance will decline, resulting in abnormal losses and fatalities. Therefore, research in Prognostics and Health Management (PHM) is of particular importance, as the correct operation of equipment is of great importance, faulty operation contributed not only to the increase of process costs, but also increased the degree of transport means consumption, unnecessarily reducing their efficiency and effectiveness [4].

Take subway mechanical equipment as example, such as subway fan vent valves, escalators, platform doors, car doors, gate machines, roller shutter doors, etc., There is a greater relationship between the health and frequency of use of equipment that has a high frequency of use, frequent inspection, and frequent usage. When an abnormality occurs in the process

of equipment operation, the state of its health will be reevaluated. Thus, a mathematical model of the action process study such as linear regression for evaluation [7] has a greater importance and value for gaining a deeper understanding of the PHM mechanism and principle of mechanical equipment.

a) There is a need to install vibration and current sensors, and data collection requires cable construction and installation, which is more time consuming and costly.

b) At this time, there are no or few signs of defects or faults, and it is difficult to pinpoint the exact principles of their emergence, development, and evolution of defects.

c) The labeling of defects and fault anomalies is more costly, and there are fewer defects and abnormal samples. High levels of noise are also a hindrance.

The purpose of this paper is to conduct three aspects of research: In the first part, we use deep learning methods to

(*) Corresponding author.
E-mail addresses: A. Junfeng (ORCID: 0000-0002-6303-131X) 21111093@bjtu.edu.cn, L. Jiqiang (ORCID: 0000-0003-1147-4327) jqliu@bjtu.edu.cn, Z. Hao (ORCID: 0000-0002-1462-1700) hao.zhen@bjtu.edu.cn, L. Mengmeng 983766829@qq.com

achieve the monitoring and classification of device openings and closings, the construction of two-stream multi-loss functions of convolutional neural networks. The second part, mainly using the method of machine vision, involves the construction of a 1DCNN model combined with the attention mechanism, analyzing the opening degree of each cycle, determining the abnormal situation according to the duration of the opening and degree of closing (hereinafter referred to as opening degree or degree of opening), and identifying whether an abnormal situation exists in every cycle based on the deep learning algorithm and categorizing the abnormality as well. In the third part, the simulation generates N action cycle data, and generates the health curve according to the health equation, determining the defect based on the trend of the curve, constructing the sliding window and the support vector machines (SVM) fusion model, notifying the operation and maintenance personnel to conduct maintenance. In the three-part study, there is a strong correlation, which was established through the construction of a deep learning algorithm to classify the opening degree, allowing the system to perceive the opening degree. Based on the curve distribution of each action cycle, it is possible to determine whether there is an abnormality, and then to determine different health values based on the type of abnormality, in order to determine whether there is a defect. Successive multiple action cycles are connected into a curve, and from the curve, the operation and maintenance personnel are reminded to perform maintenance on a timely basis. The first step is the foundation of the second part, and the second part is the foundation of the third part. They have collectively completed the three functions of condition monitoring, anomaly analysis, and defect early warning, laying the foundation for the technical research of PHM. The present paper provides a reliable basis for reducing the frequency of operation and maintenance by studying the characterization phenomenon of the equipment prior to the occurrence of faults or defects.

The aim of this paper is to combine the methods of machine vision, deep learning, and machine learning in order to conduct research on the micro-action of rotating machinery, and the contributions are summarized below.

a) This paper proposes a new framework called Rmcad that applies machine vision to the analysis and early warning of subway equipment micro-action. Rmcad stands for rotating machinery condition monitoring, anomaly detection, and defect early detection. In this study, the micro-action state detection system for rotating machinery has been introduced for the first time using image analysis. In addition, the machine vision method was used to analyze state monitoring, potential defects, and abnormalities, laying the foundation for intelligent operations and maintenance as shown in Figure 2.

b) The improvement of the network model named TSML-net. A new model is proposed, called TSML-net, which means a two-stream multi-loss function network, and a special loss function is constructed for the network. In order to detect the degree of device opening, the warmup method for learning rate concepts is employed, and the monitoring point is greater than the basicnet. When using the method of data simulation, one should construct a dataset with abnormalities, extract the Fast Fourier Transform (fft) characteristics of the data curve values and perform a classification in combination with the 1DCNN

improvement model method. The classification accuracy can reach above 99.0%.

c) New scenarios and new data. Collection of the video of the operation of the subway fan vent valve, escalators, platform doors, car doors, gate machine, roller shutter doors, used to study the micro-action of mechanical equipment, make the data of condition monitoring, anomaly detection, defect early warning, which can be used for finding anomalies or defects in subway maintenance scenarios.

2. Related work

Among the fields of PHM, which includes fault diagnosis, health management, and other research activities, such as defect early warning, defect diagnosis, fault early warning and positioning, fault diagnosis, maintenance plan, autonomous operation, and maintenance [12], etc., since there is some manpower waste in several industries at present because of the high frequency of maintenance, the concept of intelligent operation and maintenance has been proposed. It is important to improve perception, prediction, fault prevention abilities, and to fully understand the mechanism and principle of faults and defects, collect terminal perception data for analysis, such as current data, acceleration data, displacement data, lidar data, etc. Several studies have laid the foundation for defect detection, abnormal diagnosis, etc. Ruan, Y (Ruan, Yi) [17] et al. proposed a fault diagnosis and prognostic method based on the combination of a SVM and deep gated recurrent unit (DGRU) network optimized with hunter-prey optimization (HPO). An innovative method called graph-modeled singular values (GMSVs) is presented by Xin Wen [22] et al. for monitoring dynamic REB health conditions based on graph theory and SVD. With the improved YOLOv3 architecture, Yao, Z [24] et al. used the machine vision method to detect foreign bodies, which improved perception and target detection abilities. Several studies have investigated the relationship between mass wear of brake pads and brake process characteristics, and the wear of brake pads has been studied scientifically by Sawczuk W [18] et al. An HMCB network combining MSCN and BLSM is proposed by Xue Bin [23] et al. to predict RUL. To describe the drift of degenerate trajectory, Rodriguez-Picon LA [16] et al. integrated several hazard rate functions in the inverse Gaussian process. Based on the information criterion, two examples of the best-fitting model are selected. Nowakowski T [13] et al. proposed a new diagnostic algorithm for evaluating the performance of road vehicle drive shaft bearings in use, determining the upper and lower limits of diagnostic parameters, and predicting the likelihood of impending bearing wear and damage. An unsupervised monitoring method for rotary machines was proposed by Bielecki A[2] et al. This method consists of three stages: a multi-reference preliminary analysis of vibration signals, an automatic reference preliminary analysis, and a probabilistic analysis of vibration signals. It is suggested by Blair [3] that calibrated hybrid models, utilizing physics, knowledge, or data-driven methods to improve model accuracy and robustness, can be employed to quantify and account for operational uncertainty. According to Taghiyarrenanin, Z [19] et al., they analysed the performance of squirrel-cage induction motors. They obtained three phase currents and five vibration signals, used a deep conventional

classifier to improve diagnosis, and proposed a fine classifier that utilizes contrast loss to improve diagnosis. Vaiciunas G [20] et al. developed and proposed a method to detect axle box bearing defects.

Zhao L [26] et al. proposed deep learning-based remaining useful life prediction method with transformer module and random forest. Zhou, F [27] et al. studied an intelligent diagnosis method for high-speed railway turnouts based on support vector machines. As a result of this work, the fault curves can be identified automatically by computers and a significant amount of personnel and resources can be saved. Using it, turnout faults can be intelligently diagnosed. Kłosowski et al. [10], presented a novel algorithm for generating monochrome 2D images using multiple classification convolutional neural networks (CNNs) simultaneously to ensure the proper maintenance of equipment in the field of ultrasound tomography (UST). There are three parts to the methodology: feature extraction, fault detection, and fault diagnosis. In the current walking part bolt posture and quantity detection system, according to Al-Kahwati [1], a typical belt conveyor system includes component-level degradation models, estimation schemes for the remaining useful life and degradation rate, and vision-based detection of hazardous objects. Li, RZ et al. [11] proposed a method for estimating TCN in RUL, which showed excellent results.

Deep learning algorithms provide convenience for fault diagnosis, such as RegNet [15], DenseNet [9], Resnet [8]. Therefore, more algorithms can be selected for equipment fault diagnosis technology research. Some of the most popular algorithms used today are Deep Learning, Regular Artificial Neural Networks (ANNs), Nearest Neighbor algorithms (mainly KNNs), Naive Bayes, Decision Trees, and Support Vector Machines (SVMs) (Ochella et al, 2021). A number of deep learning algorithms have been employed in PHM research, including autoencoders (and their variants), restricted Boltzmann machines, deep belief networks (DBNs) and deep Boltzmann machines (DBMs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs). As well as RNNs, LSTMs and Gated Recurrent Units (GRUs) have been used in the literature for prognostic purposes. SundayOchell [14] et al. analyzed many algorithmic models including Deep Learning.

In the face of the quoted studies, many scholars use PHM for the fault diagnosis, which shows that the direction of the life prediction is more innovative, however, there are few researches on early warning. Early defect detection methods are rarely used, as are methods of building datasets of abnormal device status before installation. Thus, this article proposes PHM research using images. This paper describes a method based on video signal, studying the micro motion of rotating machinery equipment, and machine vision algorithms for defect detection and anomaly detection. It enables the perception of failure in advance, which reduces the frequency of frequent maintenance, and contributes to the development of intelligence operations research.

3. Material and methods

3.1 Framework of Rmcad

Various operating states of six types of rotating machines, presented in Table 1, were analysed.

Table 1.The different operating states of six types of rotating machinery.

The study subjects are subway fan vent valves, escalators, platform doors, car doors, gate machines, roller shutter doors. They all take the same network algorithm. Table 1 represents the micro-action screenshots of the six types of mechanical rotating machinery used in the subway system.
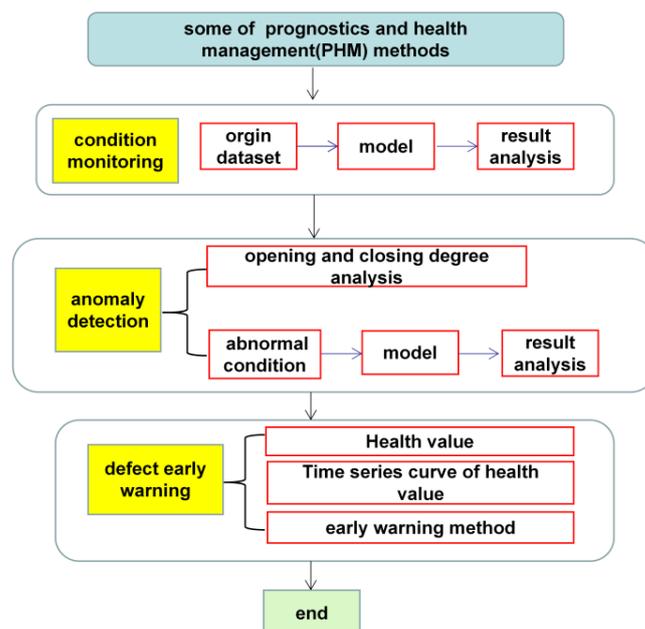

Fig. 1 Technical roadmap.

Figure 1 illustrates the research technology roadmap. The purpose of this paper is to discuss the research content of PHM of subway machinery and equipment, which contains three parts: condition monitoring, anomaly detection, and early defect detection. Condition monitoring systems monitor openings and closings of equipment, anomaly detection systems analyze abnormal data in action cycles, and defect early warning systems rely on abnormal circumstances in multiple working cycles. As a result, the equipment health curve is analysed to determine if there is a defect, which increases the ability to remotely operate and maintain the equipment as well as the technical ability of the terminal perception of the equipment, and further alarms are generated when the equipment is defective, providing theoretical support for modifying frequent maintenance cycles.

Rmcad is the name of the new framework illustrated in Figure 2. Using machine vision, the operation status of the equipment is monitored, the opening degree is calculated, and the network algorithm model is improved to improve detection accuracy. Then, the model's opening and closing values are used to calculate the abnormal opening degree in a given period, followed by the accumulation of data and construction of models to determine the abnormal action classification, further relying on the self-defined health value. Statistics are derived of abnormal conditions in each action cycle within a certain period of time, forming a time series curve of health, analyzing whether there is a defect warning situation, and early warning will be triggered under model analysis. In addition, all the equipment described above, such as fan vent valves, escalators, platform doors, car doors, gate machines, and roller shutter doors, follow these principles.

Based on Figure 2, opening degree data as well as anomaly labels and status labels are analyzed so that the camera can be considered a new sensor, because it is defined through the software serviceas a form of video sensor, which is different from the conventional current, voltage, temperature sensor. At the same time, using the new sensor perception data of opening degree, anomaly label, and health label, can greatly reduce the labor intensity of the subway operations personnel and allow more comprehensive perception of equipment status.

Also, we provide some related definitions, as shown below. Action cycle. Subway machinery and equipment perform a complete action process, such as moving from one relatively fixed state to another relatively fixed state.

Micro-action of subway rotating machinery. Rotation of subway mechanical equipment is known as micro-action in an action cycle, and micro-action abnormalities include a certain opening degree position Caton, or several opening degree position Caton. When micro-action are too frequent, mechanical equipment will be adversely affected, and its lifespan will be shortened over time.

Health of subway machinery and equipment. In the event that mechanical equipment abnormalities or defects are found, we evaluate the equipment according to the abnormalities and defects found, and we assess its overall health.

Experimental data enhancement. As a result of the noise in a real environment, some disturbances are inevitable, but at the same time, the network model is not always 100% accurate, so the interference ability of the model can be verified experimentally. In order to make the model more robust, some of the opening and closing data are adjusted to other opening and closing values after the opening degree data for a cycle are constructed.

Subway mechanical equipment opening degree. In the operation of mechanical equipment, for the equipment between a relatively fixed state A and another relatively fixed state B process, there is an intermediate transformation or action state, the state of the A, B process is divided into n parts, and one can find the process state description value of the equipment, which is the degree to which the equipment opens or closes, with a range between 0 and 1.

The opening degree of the subway fan vent valve is defined as the opening degree, the original state is when the vent valve is closed in place, the termination state position is opened in place, and the intermediate process corresponds to the corresponding rotation angle.

The degree of escalator opening is determined by the camera, which captures first the escalator, then locates the main part of the escalator, the highest definition state A, the lowest definition state B. Using the displacement of the red marker in the elevator, it maps the opening and closing of the escalator.

A platform door's and a car door's opening degrees are defined and the original state of the door is when the door is closed in place, the stop state position door is opened in place, and the value of the intermediate process is mapped to the degree of opening.

The opening degree of the gate machine is defined as the original state and represents the gate fan blade being completely closed in place, the termination state means the gate fan blade being completely opened in place, and the intermediate process opening degree value corresponds to the door opening degree.

The degree of opening of roller shutter doors is defined as follows. In order to determine the width and height of the opening and closing of the roller shutter door, it is necessary to map the original state of the door at its top, the termination state at its bottom, and the intermediate process at its intermediate state to determine the width and height of the opening and closing.
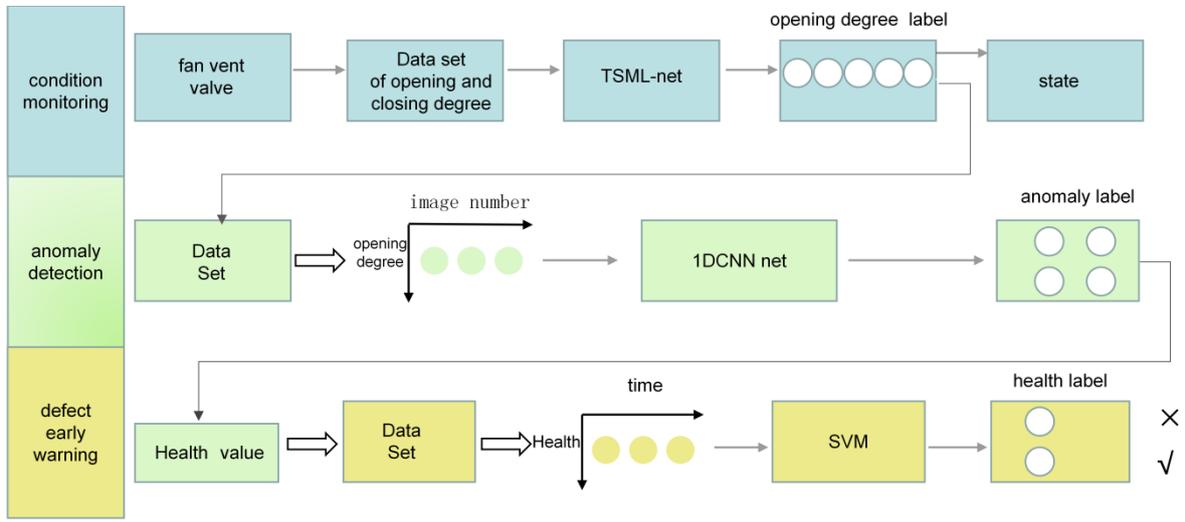
Fig. 2 Rmcad Framework (using fan vent valve as an example).

3.2 TSML-net for condition monitoring

3.2.1 Basic knowledge introduction

As part of pattern recognition, data collection, feature extraction, classification prediction, etc., is a classic topic. In the case of an escalator, for example, one can obtain the temperature of the handrail during its run as a research object, obtain the time series matrix, and then extract features from the matrix using methods such as Fast Fourier Transform (fft), Empirical Mode Decomposition (EMD), etc., and then construct a classification model based on Back Propagation (BP), SVM and other classifiers. In terms of classification models, SVM are used to solve the maximum segmentation hyperplane problem, which can be expressed as the following constraint optimization problem, which has been used by scholars in a wide range of research areas. As shown in equation (1), the hyperplane is represented by (w, b).

$$\text{Min} \frac{1}{2} \|w\| \quad s.t. \ y_i(wx_i + b) \geq 1, i = 1,2,3..N \quad (1)$$

As a popular research theory, convolutional neural networks include convolutional, pooling, full connection and other operations, which can realize local connection and weight sharing, providing the necessary foundation for classification, object detection, semantic segmentation, instance segmentation, super-resolution enhancement, and speech processing. The main introduction to convolution networks includes: the derivation of convolution, pooling and other operations, the gradient solution method, and so on. Layer l is a convolution operation, and its operation is as follows:

$$z^l = w^l \otimes m^{l-1} + b^l \quad (2)$$

Z^l the output before the activation of layer l, m^{l-1} is the output after the activation of layer l-1, $\otimes$ represents convolution, b represents bias, w represents the weight of the convolution kernel. Generally, a0 batch normalization layer can be added between the convolution kernel and the activation layer. If layer l is a fully connected operation, then the operation is:

$$z^l = w^l m^{l-1} + b^l \quad (3)$$

If layer l is a pooling operation, then the operation is:

$$z^l = f_{pool} m^{l-1} \quad (4)$$

Z^l is the output after pooling, f_{pool} is the pooling function, including maximum pooling, average pooling, Singular Value Decomposition (SVD) pooling, etc. m^{l-1} represents the output after l-1 layer.

In the case that layer l is a softmax operation, then the output is the number of categories classified, and the operation is:

$$out = softmax(z^l) = \frac{e^{z^i}}{\sum_{i=1}^{C} e^{z^l}} \quad (5)$$

out represents the output after softmax, ranging from 0 to 1, C represents the total number of categories, and z^i is the output value of i nodes.

Backpropagation parameter update algorithms include Adam, SGD, AdaGrad, AdaDelta, RMSProp, Nadam, etc. For the purpose of updating the network parameters, the gradient descent method is used.

$$w^l = w^{l-1} - lr \frac{\partial J}{\partial w^{l-1}} \quad (6)$$

$$b^l = b^{l-1} - lr \frac{\partial J}{\partial b^{l-1}} \quad (7)$$

The cost function is represented by J, the weight by w, the bias by b, and the learning rate by lr.

In the case of a subway vent valve, the current state can be transmitted to the BAS (environment and equipment monitoring system) and it can then be analysed to determine whether the vent valve is in a state of opening or closing, whether the vent valve is located in the middle position, or how much of its opening is unknown. If the vent valve is often kept in the opening and closing process, it may cause stuttering and other mechanical problems, and with the accumulation of time, this will affect the service life of the equipment, and will increase the risk of equipment failure, therefore, the value of the opening degree is of great importance.

3.2.2 TSML-net for condition monitoring

There were mainly three improvements involved in this project: the improvement of the network architecture, the improvement of the loss function, and the selection of learning rate. The comprehensive use of the best results can result in the detection and analysis of the device state and degree of opening.
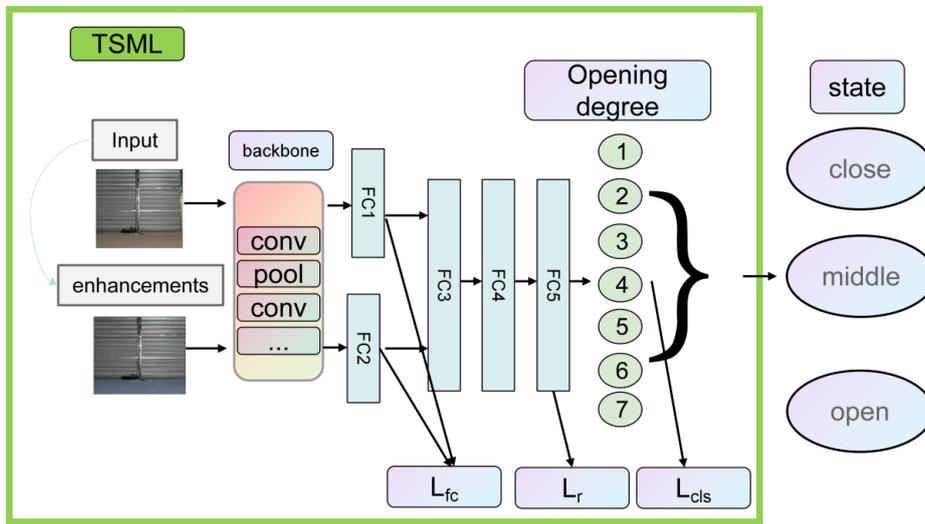
Fig. 3 Model architecture for TSML-net.

Network improvement. Figure 3 shows a TSML-net with a two-stream multi-loss function that can be used to classify openings and closings of vent valves and other subway mechanical equipment. Two-stream is the differential network, the original image is one, while the enhanced network is the other. After passing through the backbone network, the input data is connected to the fully connected layer. After the channel merger, it passes through two fully connected layers and a classification layer, where the backbone network uses the existing network, such as Lenet, Alexnet, Resnet, Googlenet, etc., by using the first convolution layer to the first full connection layer of the backbone network. Three loss functions are designed in TSML-net as shown in the figure, namely classification loss, FC layer contrast loss, and loss of intra-class and between class ratio which are referred to as L_{cls}, L_{fc}, L_r, and are back-propagated to the neural network for monitoring the degree of opening. A category label is represented by the opening, and this model represents the classification.

Process of two-stream net operation. Through weight sharing, a two-stream backbone network provides access to a concat layer. A stream using the original image is able to obtain the feature F1 through the backbone network, another stream uses weakly enhanced data from a raw image as input, and through the backbone network will reach the last full connection layer, which will enable the feature FC2 to be obtained. FC1 and FC2 can then be operated in series, and they will then be connected to the three full connection layers.

The loss functions of L_{cls}, L_{fc} and L_r. As a result of the backbone network to obtain the feature, calculating the mseloss loss function L_{fc}, then calculating the interclass ratio loss function L_r after the FC5 layer, the dimension is reduced to two dimensions, and it is not necessary to further reduce the dimension algorithm (such as Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Singular Value Decomposition (SVD), etc.). Through the use of convolutional networks, it is possible to compute the class center, to observe the visual characteristics of different classes, and to obtain the classification value of the opening degrees. In this layer which is aimed at calculating the classification loss L_{cls}, the weights of the three loss functions are of utmost importance. Different backbone networks can result in different outcomes for different data sets. As mentioned above, the FC5 is a fully connected feature layer with an output dimension of 2. It is intended to introduce a new loss function to facilitate the calculation of class centres, increasing the distance between classes, reducing the distance within the class, and improving the anti-interference ability. This will lead to a higher accuracy rate.

In addition, subway maintenance personnel need to be aware of the state of the vent valve within the subway environment, so condition monitoring is necessary. It can realize the conversion of the three states with on, off, and half open, as shown in the figure 3. If the opening degree is 1, it corresponds to a closed state, if it is 7, it corresponds to an open state, and the other degrees correspond to middle states.

Improvement of multiple loss function. The loss function of simCLR [5], Barlow Twins [25], NCE [6], Image Rotations, and Deep clustering are all classic examples of self-supervised learning and comparative learning. By using certain technical methods, the enhanced image of the original data is obtained, and this image is used as the similar sample. In order to achieve the purpose of auto-enhancement, enhanced samples of other images are used as dissimilar samples, and a model is trained to learn the differences. The two-stream is proposed to go through the convolution network, then flat to the fully connected layer, and the ratio is used as the loss function. The L_r (loss of intra-class and between class ratio) was presented in this paper. In this paper, a Two stream open degree classification method is proposed based on mechanical equipment, a given data set contains category labels $D_1 = \{(x_i, y_i)_{i=1}^{N1}\}$. In general, C is greater than 4 integers and less than 10, while y represents the open degree. In order to enhance the accuracy of classification, the raw data needs to undergo weak enhancement processing, including increasing image brightness, filtering, and adding noise. Once the changes have been made, the data set can be obtained. $D_2 = \{(x_j, y_j)_{j=1}^{N2}\}$, and then dimension of x_j, y_j is m * n, $N1$, $N2$ is the number of samples. In general, $N1=N2$ take a semi-supervised method to learn a classifier $f(x_i) \rightarrow y_i$, to make the classification accuracy as high as possible.

In this paper, three types of loss functions are used, as shown in equation (9), equation (10), and equation (11). There are three main types of loss functions, namely the classification loss

function L_{cls}, the defined loss of intra-class and between class ratio L_r, and the probability distribution loss function of two fully connected layers L_{fc} . In this paper, the loss function is computed as the weighted sum of the three variables.

$$L = a1 * L_{cls} + a2 * L_{fc} + a3 * L_r \qquad (8)$$

In equation (8), the loss L represents the total loss in the network. Hyperparameters a1, a2, and a3 are used to optimize the network model.

In equation (9), the initial loss L_{cls} is defined as a cross-entropy loss function. p and q represent the true labels and corresponding predicted probabilities.

$$L_{cls} = H(p,q) = -\sum_x p(x) * \log(q(x)) \qquad (9)$$

In equation (10), as a result of the two full connection layers, the loss function L_{fc} was obtained. We believe that the output probability distribution of the two streams should be as consistent as possible in order to resist interference. As for the stream input noise image, its classes should be the same, and adopt the mse loss function when using two full connection streams.

$$L_{fc} = loss(x_i, y_i) = \frac{1}{N}\sum_{i=1}^N (x_i - y_i)^2 \qquad (10)$$

For loss of intra-class and between class ratio L_r, it is the stream after combining FC1 and FC2. The network in this stream should be capable of clearly distinguishing categories, so there should be a greater distance between different categories, and a smaller distance between the same categories, so that the model can be more robust. In fact, the smaller the ratio between the distance between classes and the distance between classes, the better. As a general rule, the equation is as follows (11):

$$L_r = \frac{L_{inter}}{L_{between}} = \frac{\sum_{i=1}^{N1}(g_i - c_{y_i})^2}{\sum_{i=1, j=i}^{C}(c_i - c_j)^2 + \varepsilon} \qquad (11)$$

ε is an infinitesimal decimal, where i and j are integers lesser than the total number of classes C, where the center point of the training samples must be updated by mini-batch processing, c_i, c_j represents the center of the class corresponding to the 2D features of the samples, $y_i \in \{1,2..C\}$, and C represents the total number of open and close classes. An analysis of two-dimensional features obtained after passing i samples of g_i representative through the backbone network, $g_i \in R^n$. The smaller the L_r, the better. The equation for updating the in-class center is $c_j^{t+1} = c_j^t - lr * \Delta c_j$, c_i is the in-class center, lr is the learning rate, and the solution equation of the class center is in equation (12).

$$d_i = [x_i, y_i], c_i = \left[\frac{\sum_{i=1}^{N1} x_i}{N1}, \frac{\sum_{i=1}^{N1} y_i}{N1}\right] \qquad (12)$$

where d_i is the characteristic value of dimension 2 and $N1$ is the number of samples.

In the following equation (13), the back propagation method is used to update the network parameter θ , x_i represents the input value for the corresponding layer in the network.

$$\theta_{t+1} = \theta_t - lr * \frac{\partial(a1*L_{cls} + a2*L_{fc} + a3*L_r)}{\partial x_i} \qquad (13)$$

The selection of the appropriate optimizer and learning rate warmup method. Optimizers and learning rates are used. The optimizer adopts the strategy of AdaBound, and the learning rate adopts the strategy of warmup and cosine annealing. Dynamic learning rate boundaries are applied to achieve a smooth transition from adaptive methods to SGD, as well as to provide a theoretical demonstration of convergence. Experimental results indicate that the new variant can eliminate the generalization gap between the adaptive method and SGD, while maintaining a high learning speed during early training, as well as ensuring strong generalization abilities on test data. Back propagation primarily completes parameter updates in neural networks:

$$\theta_t = \theta_{t-1} - lr * g_t \qquad (14)$$

g_t is the gradient update quantity, and warmup is the method used to adjust the learning rate. ResNet model describes a method for warming up the learning rate.
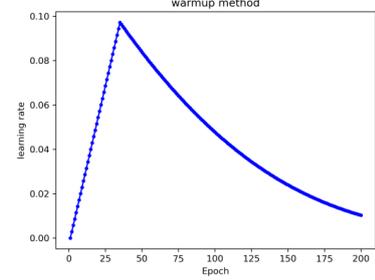

Fig. 4 Learning rate curve taken under warmup method.

Figure 4 illustrates the learning rate. Learning rate does not stay constant as training iterations increase, but constantly changes. We trained some epochs, then modified the learning rate to the preset rate. Before 35 steps, learning rate increases gradually, and after 35 steps, it declines. Table 2 presents the Algorithmic Steps of the Model for condition monitoring.

Table 2. Algorithmic Steps of the Model for condition monitoring.

condition monitoring steps
Begin in: 1. Pooling of the required data 2. The video is divided into frames to obtain the labels Y1 and Y2 for each frame, where Y1 represents the status label of the device and Y2 represents the value of the opening . 3. The image is taken as the input of the neural network, and the label is taken as the output of the neural network to construct the deep learning algorithm model, and the model effect is viewed from the loss rate. 4. A two-stream network is constructed. One stream performs data analysis on the original image to obtain FC1, and the other stream performs data analysis on the weak enhancement image, and then obtains FC2. 5. FC1 and FC2 are combined, and then the FC layer is linked to obtain the final output value. 6. Optimization of parameters, equation of the relevant loss function, and defining the learning curve Input: for labeled raw data and weakly enhanced datasets. $D_1 = \{(x_i, y_i)_{i=1}^{N1}\}$, $D_2 = \{(x_j, y_j)_{j=1}^{N2}\}$, Initializing the parameter θ, the hyperparameter, a1, a2, a3, the learning rate lr and the maximum number of iterations, epoch Output: parameters θ . Output Indicates the parameter 6.1 epch=1 6.2 repeat 6.3 The value of the loss function is calculated by the equation L_{cls}, L_{fc}, L_rloss function, and total loss function 6.4 for i < N1, j < N2, do as the equation (15). So $\frac{\partial L}{\partial x_i}$can be solved 6.5 Updating the class center c_i. 6.6 Updating parameters θ via equation(13) 6.7 epch=epch+1 6.8 while epch<epoch, continue 7 Training of the model and the parameter tuning 7.1 Improving the optimizer 7.2 Improved lr iteration method 8 Optimization model improves the effect of the algorithm model。 end

$$\frac{\partial L}{\partial x_i} = a1 * \frac{\partial L_{cls}}{\partial x_i} + a2 * \frac{\partial L_{fc}}{\partial x_i} + a3 * \frac{\partial L_r}{\partial x_i} \quad (15)$$

After reducing the dimension of the network design, a fully connected layer can obtain two-dimensional features, and the data model differentiation effect can be visualized.

3.3 1DCNN for anomaly detection

The main method. One should simulate a large number of abnormal data, apply the data enhancement method to increase the image data of the intermediate state, and then create a new video data set based on the simulation to produce different opening degree state data that is then merged with comprehensive data sets. For the data generated by the simulation, the value of the mechanical opening degree can be obtained for each frame of the video, and then all frame images are recorded as a row vector A, and all the opening and closing data is obtained, as well as the training sample data. For the training sample, the network model based on 1DCNN is used, and the abnormality type is determined, and the detection effect of the abnormality type is demonstrated. As a result of studying the main anomalies of the vent valve, such as caton and lack of oil, it is evident that the occurrence of these anomalies will not affect the normal operation of the mechanical equipment. In this paper, the ECA attention mechanism [21] model's mind is primarily used to design the improved architecture of 1DCNN, and abnormalities are recognized.
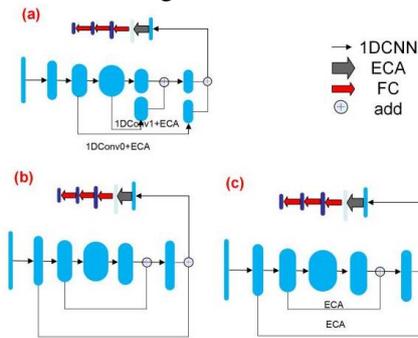

Fig. 5 Model architecture for 1DCNN(a), 1DCNN(b), 1DCNN(c).

We have designed the three different architectures. Improved network architecture for anomaly detection design is presented in figure 5. Let us take the 1DCNN (a) as an example. It illustrates the original input matrix, which is a one-dimensional row vector matrix connected to six 1DCNN modules whose output channel sizes are 32-64-128-64-32-1, forming an inverted shape, details can be found in Table 3. Next, the ECA attention mechanism module is connected to three fully connected layers that enable classification of opening and closing categorical anomalies.
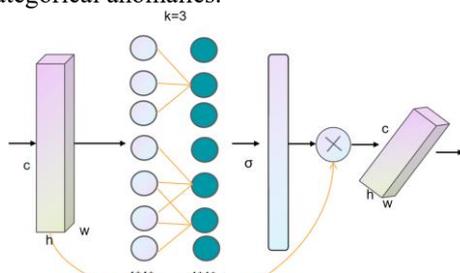

Fig. 6 Diagram of the ECA attention mechanism module.

Figure 6 shows the schematic diagram for the ECA Attention Mechanism Module. Based on global averaging, ECA performs a fast one-dimensional convolution of size k, where k is adaptively determined by mapping through stream dimension C. As this is a vector of one dimension, w means width of input data and it is 1, whereas the value of k is 3.

```python
class eca_layer(nn.Module):
    def __init__(self, channel, k_size=3):
        super(eca_layer, self).__init__()
        self.avg_pool = nn.AdaptiveAvgPool2d(1)
        self.conv = nn.Conv1d(1, 1, kernel_size=k_size,
                              padding=(k_size - 1) // 2, bias=False)
        self.sigmoid = nn.Sigmoid()
    def forward(self, x):
        # x: input features with shape [b, c, h]
        x = torch.unsqueeze(x, 3)#new code
        b, c, h, w = x.size() #
        y = self.avg_pool(x)
        y = self.conv(y.squeeze(-1).transpose(-1, -2))
        y=y.transpose(-1, -2).unsqueeze(-1)
        y = self.sigmoid(y)
        return x * y.expand_as(x)
```

Fig.7 Code written in PyTorch.

In figure 7, the main code using PyTorch is shown. Since x is a vector of one dimension, we must use the unsqueeze function to expand the original data to four dimensions. In the figure this means adding one dimension after the third dimension. One should follow the ECA algorithm process to implement the rest of the code.

Considering the layers named $conv1d$, $relu$, and $maxplling$, the information about the network model in figure 5 is provided in Table 3. Among the information displayed are the input size, operator, parameter and outsize. Of course, in the ECA module connected after 1Dconv0, the resulting data is 4 dimensions, so the last dimension needs to be removed before the subsequent operation.

Table 3. Overall architecture of 1DCNN(a) network model.

no	Operator	parameter	Param #
1	Conv1d, relu, maxplloing	Conv1d:32, k=10, s=1, p=0 Maxplloing:k=2, s=1, p=0	352
2	Conv1d, relu, maxplloing	Conv1d:64, k=3, s=1, p=0 Maxplloing:k=2, s=1, p=0	352
3	Conv1d, relu, maxplloing	Conv1d:128, k=3, s=1, p=0 Maxplloing:k=2, s=1, p=0	6208
4	1Dconv0, relu, maxplloing,ECA-attention	Conv1d:32, k=3, s=1, p=0 Maxplloing:k=2, s=1, p=0 ECA: k=3	18, 464
5	Conv1d, relu, maxplloing	Conv1d:64, k=3, s=1, p=0 Maxplloing:k=2, s=1, p=0	24,640
6	1Dconv1, relu, maxplloing,ECA-attention	Conv1d:64, k=3, s=1, p=0 Maxplloing:k=2, s=1, p=0 ECA: k=3	18,464
7	Conv1d, relu, maxplloing	Conv1d:32, k=3, s=1, p=0 Maxplloing:k=2, s=1, p=0	24,640
8	Conv1d, relu, maxplloing	Conv1d:1, k=3, s=1, p=0 Maxplloing:k=2, s=1, p=0	6,176
9	ECA-attention	k=3	97
10	linear	275-128	35,328
11	linear	128-32	4128
12	linear	32-4	132

The research and field investigation indicate that the subway equipment and machinery will display caton, lack of lubrication equipment, repeated action, equipment failure, mechanical equipment damage, as well as other issues.
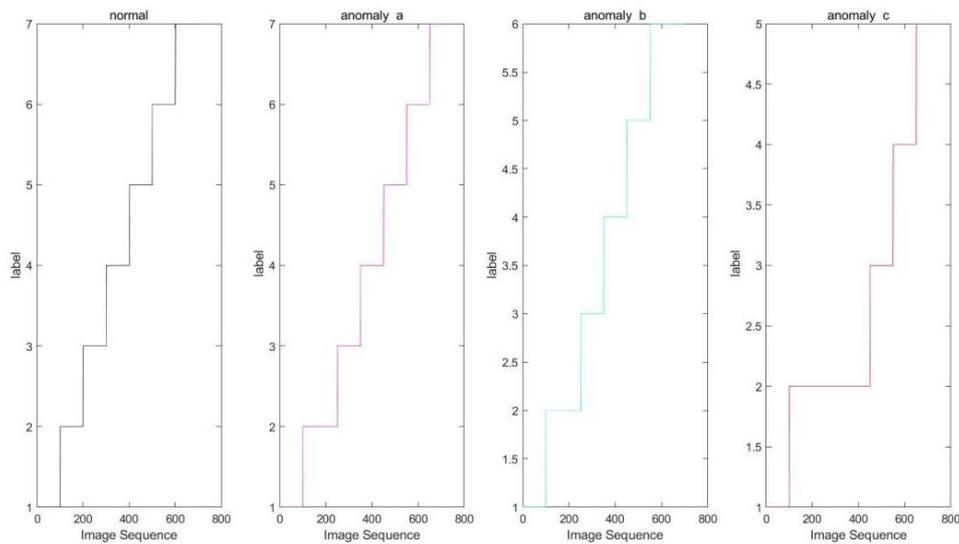
Fig. 8 Examples of samples for anomaly detection.

As shown in fig. 8, this paper describes three types of anomalies: a, b, and c. The abnormal a represents the number of frames exceeding n frames, the abnormal b represents the number of frames exceeding n (suppose the number of normal state frames is ss, $ss < n$) dimensions in two open and closing positions, and the abnormal c represents the maintenance number of frames exceeding m in some opening and closing positions. As shown in figure 8 above, the label means the opening degree category label, it shows the four different samples. This design anomaly detection data set, coupled with algorithm research, the main algorithm steps are shown in Table 4.

Table 4. Anomaly detection algorithm steps.

anomaly detection steps
begin
1 The raw video is obtained
2 Data augmentation of the images obtained from the video
3 A new video is obtained from the expanded image combination
4 The opening degree detection method is used to obtain the row vector
5 Combined with simulated data, the training sample A is produced
6 Using sample A, train and test the model of 1DCNN
end

By following the above process, anomalies were detected under different action cycles, which is helpful for analyzing the frequency of abnormal actions and potential defects in the future. After accumulating the opening degree data, several data classification methods can be applied, mainly including the following: a) drawing a curve based on the original data, and then saving that curve as a JPG image, and then realizing the classification of the image, b) the original graphic is extracted using features such as EMD, fft, etc., and then a classification study is conducted using machine learning, c) as the original data is changed, for example by using the wavelet time-frequency map to obtain the changed image, the image classification operation is carried out. Therefore, when designing the scheme, many methods can be adopted, including, but not limited to, EMD+BP, EMD+SVM, EMD+Random Forest (RF), fft+1DCNN, Alexnet image classification, resnet classification, wavelet time frequency + alexnet. This paper primarily uses fft method for feature extraction, followed by the 1DCNN network algorithm for model classification.

3.4 Defect early warning

To simulate the occurrence and evolution of defects, defect data are created. The Monte Carlo data enhancement method and theory are applied. As a result of the present frequency of maintenance, the phenomenon of maintenance often occurs, as are equipment failures and micro action potential abnormalities at the same time. Too much potential risk can result in equipment failure, which may affect the equipment's service life over time. To diagnose potential risks and potential abnormalities, Machine Learning combined with video surveillance is adopted to improve the service life and reduce equipment failure risks. In order to increase the service life of the equipment, the Monte Carlo method is used to simulate the health value of N action cycles.

Calculation of the health value. The health value of the equipment is determined based on its use and the number of abnormal and defective times $y = 100 - y1 - y2 - y3$. As $y1$, $y2$, $y3$ represent the use of the situation and abnormal deduction points, obviously, the higher the health value, the better.

Design of defects. On the basis of the T1-T2 time series, as in equation (16), this paper defines three defects: (a), (b), (c), where defect (a) indicates that health maintains a low value of p1, defect (b) mainly depends on slope k. If g (x) slope k is less than 0, then there is a defect. Defect (c) represents the number of values smaller than threshold over a certain number m. The defect (a) represents the score in the same low score for n cycles, it needs to be recalled to be examined, in a continuous sliding window represented by the defect (b), the score data continued to decline. As can be seen, the derivative value of this part is a large negative number. The equipment performance could continue to decline if the operator and maintenance personnel do not pay attention. Defect (c) represents a sliding window, it scores less than the threshold q1 exceeded m times, (q1 can be set to 50, m can be set to 8). As a result, the operation and maintenance personnel should be reminded to check the site. In addition, this method can provide ideas for predictive maintenance. In order to achieve intelligent operation and maintenance, the original frequency of frequent maintenance can be greatly reduced. Here is the equation for the three defects.

(a): $f(x) = p1, p1 < \mu, \mu > 0$
(b): $g(x) = -kx, k > b1, b1 > 0$
(c): $sum(f(x) < q1) > m, \quad q1 > 0$ $\quad\quad\quad$ (16)

Methods for designing algorithms and mechanisms for detecting early warning signs. SVM and sliding window are the main ideas, as can be seen from the Table 5. Using a sliding window, we collect data on the number of abnormal behaviours during each action cycle, health values for each certain sliding window, then label each sample, and then construct an SVM classification network, based on the SVM judgment results to diagnose a defect. If a defect is found, the appropriate operation and maintenance personnel will be notified to attend the maintenance, there is no need to rush to the scene frequently if there is no defect.

Table 5. Model steps for defect detection.

detecting defects steps
Begin with:
1 According to the previous chapter model, each action cycle's abnormal behavior is counted and recorded
2 The index values are solved according to the device health definition, and the data are placed in the time series matrix
3 The index value time series curve is recorded after some accumulated statistics have been accumulated
4 Time series curve analysis using sliding windows and SVM algorithms
5 Determining whether there are any defects in the SVM model based on the results
6 If a defect is found, inform about it, otherwise, return to step 4
end

4. Results

4.1. Experimental Environment

In this experiment, experiments involving anomaly detection and condition monitoring have adopted the Python programming language, video enhancement has adopted the m language, and defect warning has adopted the m language. The detailed characteristics of the computer used in this experiment are described in Table 6.

Table 6. Basic machine characteristics.

Hardware and software	Characteristics
Memory	8 GB
Processor	11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz 2.30 GHz
Graphics	NVIDIA GeForce RTX 2080
Operating system	Win10, Python 3.9, PyTorch 1.80, matlab2016

4.2 Effectiveness experiment of state monitoring

The data set and parameters of the network. Rotating machinery

equipment includes subway fan vent valves, escalators, platform doors, car doors, gate machines, roller shutter doors, As expected, the results of these devices will be consistent, and obtaining the opening degree will be possible by marking the fixed position of the escalator with special markings. Their degree labels were 7, 6, 4, 4, 4, 5, respectively, and more than 3, 000 images are sampled for each device. As part of the field collection and network collection, we observed the entire equipment continuously, without interference from personnel or objects, collected data from the field collection and network collection, and collected network equipment operation videos from China's Jinan Subway (a total of 100 videos) and then cut the videos to obtain a number of effective videos. Approximately three months were required to collect the data, including the collection of subway fan vent valves, escalators, platform doors, car doors, gate machines, roller shutter doors action videos. In this paper, the original image of the fan vent valve dataset is divided into seven labels based on the size of the opening degree, and other equipment is used to verify the validity of the algorithm model. The rotation angle of the fan vent valve is $0°, 15°, 30°, 45°, 60°, 75°, 90°$ Semi-open data concerns those with an opening angle of $15°$ to $75°$. If we take the fan vent valve as an example, collect video data, select the effective part, convert into images, and divide into 7 folders representing 1 to 7, most of the datasets are above 3,000 samples. Multiple sets of data were constructed for training and learning. The origin data is enhanced weekly and as the second channel of the TSML-net, then the training and test samples are split 8:2. As described in Section 3.2, network parameters were selected.

Basic net. It has two convolutional layers of construction, of which the number of convolutional kernels is four and six, and the size of convolutional kernels is three. The accuracy values of the benchmark algorithm are displayed in Table 7, where acc means accuracy rate.

Table 7. Benchmark algorithm accuracy values.

benchmark algorithm	train-acc	test-acc
basic net	92.3%	91.7%
TSML-net only use L_{cls}	94.2%	93.2%
TSML-net $+L_{cls}+L_{fc}+L_r$	95.6%	95.2%
TSML-net $+L_{cls}+L_{fc}+L_r$ +warmup learning rate	95.8%	95.5%

TSML-net can be seen in figure 2, while the warmup learning rate according to the proposed method is presented in figure 4. Table 7 presents the results for various algorithms by using TSML-net, multiple losses, and a warmup learning rate, for the training and testing set.

Table 8. Comparison of performance under different backbone networks on different experiments.

backbone		none	$+L_{cls}$	$+L_{cls}+L_{fc}$	$+L_{cls}+L_{fc}+L_r$
Alexnet	train-acc	96.8%$\pm$ 0.1%	97.7%$\pm$ 0.5%	99.3%$\pm$ 0.2%	99.4%$\pm$ 0.2%
	test-acc	96.3%$\pm$ 0.1%	96.5%$\pm$ 0.2%	98.3%$\pm$ 0.3%	98.4%$\pm$ 0.2%
mobilenet	train-acc	97.9%$\pm$ 0.1%	97.8%$\pm$ 0.3%	99.3%$\pm$ 0.1%	99.5%$\pm$ 0.2%
	test-acc	96.8%$\pm$ 0.1%	96.9%$\pm$ 0.3%	98.4%$\pm$ 0.1%	98.5%$\pm$ 0.1%
Lenet	train-acc	95.1%$\pm$ 0.1%	95.5%$\pm$ 0.4%	98.8%$\pm$ 0.1%	98.9%$\pm$ 0.2%
	test-acc	94.4%$\pm$ 0.2%	94.5%$\pm$ 0.4%	97.2%$\pm$ 0.3%	97.3%$\pm$ 0.2%

In Table 8, the accuracy of the test set will be about 1.5 percentage points lower than the accuracy of the training set. None means no two-stream network is used, $+L_{cls}$ means using TSML-net architecture and using L_{cls} as the loss function, $+L_{cls}+L_{fc}$ means using TSML-net architecture and using $L_{cls}+L_{fc}$ as the loss function, $+L_{cls}+L_{fc}+L_r$ means using TSML-net, $L_{cls}+L_{fc}+L_r$ as the loss function. Through the verification of the three backbone networks, it can be seen that the combined effect of the three loss functions is more effective than the effect of one loss function alone. The combination of the three loss functions will be more robust.

Table 9. Comparison of performance under different hyperparameters.

backbone	picture size	a1, a2, a3	plies	acc
Alexnet	227	1, 1, 0.2	28	99.4%
mobilenet	224	0.1, 0.7, 0.3	269	99.5%

In Table 9, picture size represents the input data size, a1, a2, a3 represent the weight of the loss function, plies represents the number of layers of the network, and parameters represents the number of parameters in the network.

The Alexnet network connects 8192 features through the backbone network, the mobilenet network connects 4096 features through the backbone network to obtain the two-stream image concatenation, and then the two networks are connected by four fully connected layers, self.fc1 = nn. Linear(8192, 120), self.fc2 = nn. Linear(120, 84), self.fc3 = nn. Linear(84, 2), self.fc4 = nn. Linear(2, 7). Using different values of hyperparameters a1, a2, and a3 gives completely different results. As shown in Table 9, the hyperparameters can be adjusted to achieve a good accuracy.

Table 10. Values of accuracy with different weak enhancements.

enhancements	acc
Random noise	99.5% $\pm$ 0.3%
Contrast enhancement	99.2% $\pm$ 0.2%
Gaussian noise	99.1% $\pm$ 0.1%
Luminance enhancement	99.5% $\pm$ 0.1%

Table 10 shows the accuracy of different data enhancement methods, and all reach more than 99.0%, proving the algorithm's robustness.
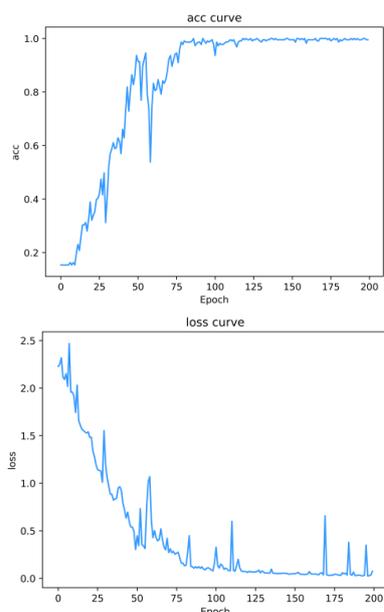

Fig. 9 Results of the accuracy rate and loss curve visualization.

The accuracy rate and loss iteration curve values are shown in figure 9, and as the number of iterations increases, the accuracy rate improves and a better classification effect is obtained. It can be seen from the trend of the curve that the algorithm model is effective.
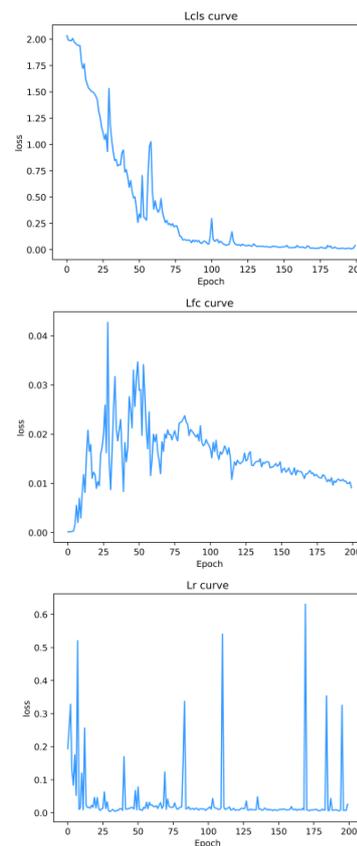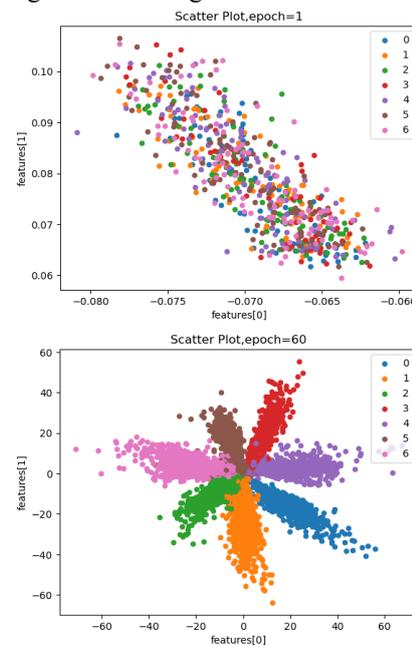

Fig. 10 Loss function iteration curves.

According to figure 10, with an increase in iterations, the L_{fc} loss function decreases, the L_{fc} loss function value is low, the L_r iteration curve jumps, and the overall loss function value is small, allowing different categories of data to be separated.
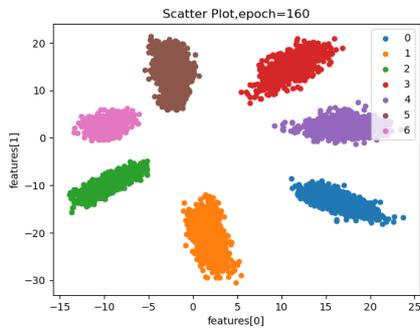
Fig. 11 category visualization under different numbers of iterations.

According to figure 11, after the visualization of the data with 1, 60, 160 iterations, the images with different degrees of opening are easier to classify at 160 iterations. In order to get the final classification result, the corresponding full connection layer must access the classification layer (opening degree). As shown in Figure 11, the visualization does not require a dimension reduction technique (such as PCA, SVD, etc.). The L_r loss function detects data with FC of 2 layers, assuming [x1, x2] FC features, features [0] in the graph is the abscissa x1 and features [1] is the abscissa x2.

The chapter on Improvement of multiple loss functions describes three loss functions, L_{cls}, L_{fc}, L_r, and we can change the type of loss function, like L_{fc}, which can take mse, KL, or MMD. The loss function used in this paper has good effect, as shown in Table 11. Three loss functions are also compared, as shown in Table 11, to verify its validity.

Table 11. Loss function effectiveness analysis.

Combination type of loss function	train-acc
Loss=L_{cls}+L_{fc}+L_r, take mse loss function as L_{fc}	99.4%$\pm$ 0.2%
Loss=L_{cls}+L_{fc}+L_r, KL function as L_{fc}	98.9%$\pm$ 0.3%
Loss=L_{cls}+L_{fc}+L_r, MMD function as L_{fc}	99.1%$\pm$ 0.4%
Loss=L_{cls}+L_{fc}+L_r, centerloss function as L_r	99.1%$\pm$ 0.2%

Below are the equations for the symmetric KL divergence and the MMD.

$$KL(x, y) = \sum_{i=1}^{N} x_i * log(\frac{x_i}{y_i}) + \sum_{i=1}^{N} y_i * log(\frac{y_i}{x_i}) \quad (17)$$

$$MMD(x, y) = \left\| \frac{1}{n} * \sum_{i=1}^{n} \emptyset(x_i) - \frac{1}{m} * \sum_{i=1}^{m} \emptyset(y_i) \right\|_H^2 \quad (18)$$

H indicates that this distance is measured by mapping the data to the regenerative kernel Hilbert space (RKHS).

The main results of the study of other datasets are presented in Table 12 in order to verify the effectiveness of the experiments and the reliability of the model.

Table 12. result of other training datasets.

datasets	acc
escalators	99.1%$\pm$ 0.3%
roller shutter doors	99.3%$\pm$ 0.1%
car doors	99.4%$\pm$ 0.1%
platform doors	99.3%$\pm$ 0.2%
gate machine	99.1%$\pm$ 0.3%

By optimizing the algorithm in the Table 12, the accuracy rate increases by at least 8%, providing the foundation for subsequent anomaly detection and defect early warning analysis of mechanical equipment. The detection effect of the opening and closing state is good, the condition monitoring and opening degree monitoring functions are implemented.

4.3 Effectiveness experiments for anomaly detection

This paper uses two methods to synthesize the production of the vent valve dataset, one is to repeat the video frames N times in an action cycle, which simulates stuttering effectively. The method described in the previous section is used to obtain the degree of opening. The other one is based on the number of images in a cycle, simulating the degree of opening and uses two algorithms to create a dataset of different devices. There are 2000 samples in each dataset, 50% for training, 50% for testing. In order to increase the anti-interference power of the model, sixty samples were randomly selected to add noise. Three categories of anomalies exist in this anomaly: one degree of opening position stuttering, two degrees of opening and closing position short stuttering, long-term stuttering at a certain degree of opening, plus a normal state. Figure 9 in 3.3 chapters illustrates the four types of data.

This paper using the warmup learning rate update strategy, using the AdaBound optimizer, and then the SVM, 1DCNN and other classifier technology research.

Table 13. Different approaches and their results in different experiments.

Method	train-acc	test-ac
fft+SVM,kernel='poly'	69.4%$\pm$ 0.6%	68.4%$\pm$ 0.6%
fft+SVM,kernel='rbf'	91.4%$\pm$ 0.2%	91.4%$\pm$ 0.2%
fft+SVM,kernel='linear'	73.4%$\pm$ 0.2%	73.2%$\pm$ 0.2%
fft+1DCNN+ECA(a)	99.7%$\pm$ 0.3%	99.5%$\pm$ 0.2%
fft+1DCNN+ECA(b)	99.7%$\pm$ 0.3%	99.5%$\pm$ 0.2%
fft+1DCNN+ECA(c)	99.8%$\pm$ 0.3%	99.5%$\pm$ 0.2%
fft+xgboost	98.7%$\pm$ 0.3%	98.7%$\pm$ 0.2%
fft+KNN	72.3%$\pm$ 0.3%	71.3%$\pm$ 0.2%

Table 13 shows that the method of fft+1DCNN+ECA can be used for monitoring of abnormal conditions and effectively realize the accuracy curve of various early warning and prediction methods.
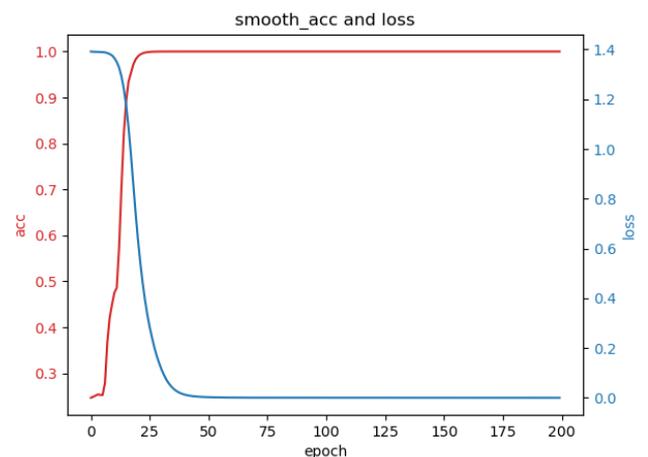

Fig. 12 accuracy curve and loss curve of 1DCNN.

The accuracy curve and loss curve are shown in figure 12. According to the accuracy rate, the method designed in this paper predicts anomalies very well and lays a good foundation for anomaly detection by having a higher accuracy rate and smoother error after 30 iterations. As a result, when we changed the addition method to the multiplication method in the experiment, we could only obtain an accuracy of 85%, proving that the network architecture shown in figure 5 is effective.

Table 14 presents the main results of the study of other

datasets in order to verify the effectiveness of the experiments and the reliability of the model.

Table 14. Results of classification for other datasets.

escalators	platform doors	roller shutter doors	car doors	gate machine
99.4% ± 0.2%	99.5% ± 0.1%	99.4% ± 0.2%	99.3% ± 0.3%	99.3% ± 0.2%

4.4 Effectiveness experiments for defect warnings

Data generation using Markov Chain Monte Carlo (MCMC). A network sample is formed by randomly selecting a certain number of samples from large sample libraries. The following examples are showing defects. The goal is to generate normally distributed data (x1, x2) with a random value of 0, 0, variance of 0.9, 0.9, and maximum minimum range of -5, 5. A sample set of 10, 000 iterations is chosen, some samples are randomly sampled for experimental data, according to the size of the collected data x1 mapping to the real sample curve position (see Figure 13). For example, if MCMC obtains data equal to 0, then the 200 position is set as the begining of a defect, and so on, to obtain the final training data.
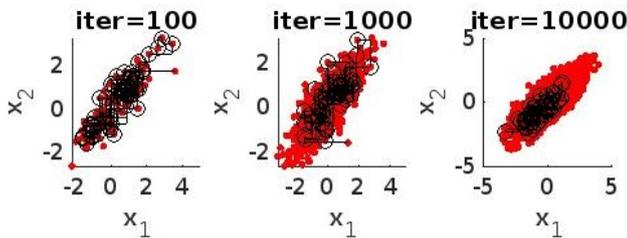

Fig 13 Data under different MCMC parameters of -5, 5.

MCMC is configured with 100, 1000, 10, 000 iterations, as shown in Figure 13, where red represents all the sample libraries, o represents the randomly sampled data used to compose the network samples. The data range appears in the range of -5 to 5 as the number of iterations increases.
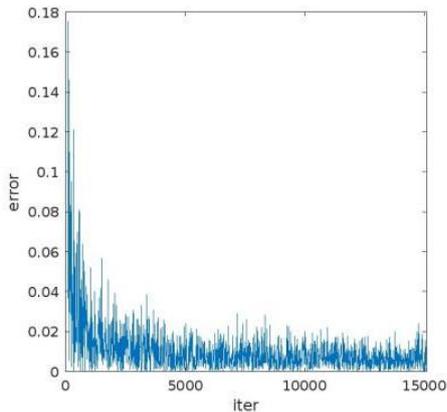

Fig 14 Variance error.

In Figure 14, the iterations were increased to 15000. When the number of iterations increases, the ordinate draws the difference between the variance of the generated data and the target variance (0.9). It can be clearly seen from the figure that the error shows a gradually decreasing trend, which indicates that the closer the sampled data is to the expected target, the closer the variance will be to the expected target as the number of iterations increases. So data generated based on 15, 000 iterations (with an error of 0.007) were selected.

Taking the vent valve as an example, a stutter at opening and

closing is worth 10 points, a temporary stutter at opening and closing is worth 20 points, and a long-term stutter at a certain opening degree is worth 30 points. Using the different abnormal working conditions of the previous chapter, the final scoring equation for equipment health is: source = 100-y1-y2-y3. During the experiment, the simulation generates data for 400 action cycles, and the abnormal data is obtained using the Monte Carlo method for each switching cycle. Figure 15 shows the health curve and health degree of each action cycle.
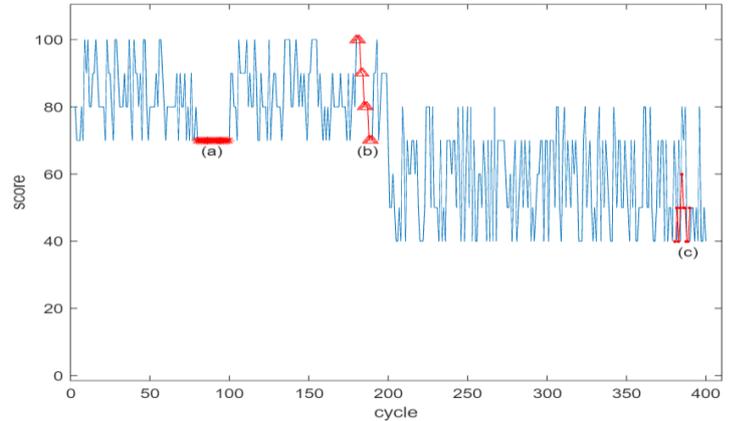

Fig. 15 a comparison of scores and defect markers across multiple cycles.

According to figure 15, the simulation generated 400 action cycles and marked three types of defects. Figure 15 clearly shows the places - (a), (b), (c) where personnel should be warned and where the need to make sure the equipment is in good working order should be indicated. This will help improve the life of the equipment.. Different domain distributions are simulated in this paper, resulting in abnormal situations in the cycle, followed by technical research, in which a sliding window + SVM method is adopted to classify abnormal categories, with a window size of 10 and 40 samples obtained, allowing the diagnosis of abnormalities and the adoption of various alarming strategies. Lastly, the simulation produces different data, and under the SVM algorithm, the accuracy rate is 100%, with parameters set to C=2, C=3, and Radial Basis Function (RBF) as the kernel function.

5. Discussion and Conclusions

This paper utilizes the machine vision method in the study of PHM management of subway mechanical equipment. The opening and closing of mechanical equipment is analysed, a two-stream model is integrated, and an interclass loss ratio function is proposed for each class. For abnormal states, the constructed 1DCNN model shows a greater improvement than SVM and KNN algorithms when combined with the proposed architecture method. The health definition for defect warning is defined, and the effectiveness of the algorithm is well demonstrated by the SVM model. In terms of state monitoring: firstly, the loss function design method of the two-stream network is given, the concept of inter-class ratio value and the concept of multi-loss function are presented, and the loss function is derived, the objective function of the model is defined, and finally, the loss function L_{fc} assumed from the FC layer is proposed, based on the concept of semi-supervised learning. It is proven that the improved CNN network using two-stream multi-loss can effectively be classified for the degree of

opening by determining its interclass ratio loss L_r, which is verified by different mechanical equipment data sets. With excellent anti-interference capability, the improved TSML-net model of two-stream multi-loss allows for double-classification of opening degrees and states, the accuracy rate can be improved over the single-stream single-loss model. The algorithm model needs to be further studied in conjunction with the current data to examine the mechanism and principle of mechanical equipment failure, and to lay the foundation for future algorithm research.

Future subway equipment needs to develop so that perception intelligence and knowledge intelligence can be utilized in the future. This paper presents a method for utilizing opening and closing data of subway equipment to diagnose abnormalities such as local catson, while combining this with the health function. A simple condition monitoring study, detection of anomalies, early warning of defects, prediction of future failures, diagnosis of defects, and prediction of

abnormalities allows the development of a maintenance plan, implementation of active operation and maintenance, and prediction of remaining operating time. This will be achieved by combining the abnormal point of defect detection with the CNN network and laying the foundation for the realization of PHM research on subways.

There are some defects in the model presented in this paper, including the construction of a two-stream multi-loss function of the network. The calculation complexity is relatively high, the need to search for other excellent algorithms. Furthermore, in the construction of a diagnostic algorithm model for detecting abnormal defects, as an alternative, one may consider the use of hybrid network models, including 1DCNN combined with LSTM algorithm networks, to construct defect detection networks, relying on the size of the sliding window for non-random data, and exploring diagnostic methods for variable sliding window sizes in the future.

References

1. Al-Kahwati, K., Birk, W., Nilsfors, E. F., and Nilsen, R. Experiences of a digital twin based predictive maintenance solution for belt conveyor systems. In PHM Society European Conference, volume 7, pages 1–8, 2022. http://dx.doi.org/10.36001/phme.2022.v7i1.3355.
2. Bielecki, A., Bielecka, M., Jablonski, A., and Staszewski, W. Simple method of failure detection of rotary machines. Diagnostyka, 22, 2021. http://dx.doi.org/10.29354/diag/142862.
3. Blair, J., Stephen, B., Brown, B., Forbes, A., and Mcarthur, S. Hybrid fault prognostics for nuclear applications: addressing rotating plant model uncertainty. In PHM Society European Conference, volume 7, pages 58–67, 2022. http://dx.doi.org/10.36001/phme.2022.v7i1.3321.
4. Borucka A. Three-state Markov model of using transport means. Business Logistics In Modern Management 2018:. 3-19.
5. Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. In International Conference on Machine Learning, pages 1597–1607. PMLR, 202.https://doi.org/10.48550/arXiv.2002.05709
6. Dyer, C. Notes on noise contrastive estimation and negative sampling. arXiv preprint arXiv:1410.8251, 2014, https://doi.org/10.48550/arXiv.1410.8251
7. Grzelak M, Borucka A, Guzanek P. Application of linear regression for evaluation of production processes effectiveness. In International Conference Innovation in Engineering 2021: 36-47, Springer, Cham, https://doi.org/10.1007/978-3-030-78170-5_4
8. He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016. http://dx.doi.org/ 10.1109/CVPR.2016.90.
9. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4700–4708, 2017. http://dx.doi.org/10.1109/CVPR.2017.243.
10. Klosowski, G., Rymarczyk, T., Kania, K., Swic, A., and Cieplak, T. Maintenance of industrial reactors supported by deep learning driven ultrasound tomography. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 22(1):138–147, 2020. http://dx.doi.org/10.17531/ein.2020.1.16.
11. Li, R., Chu, Z., Jin, W., Wang, Y., and Hu, X. Temporal convolutional network based regression approach for estimation of remaining useful life. In 2021 IEEE International Conference on Prognostics and Health Management, pages 1–10. IEEE, 2021. http://dx.doi.org/10.1109/ICPHM51084.2021.9486528.
12. Mahboub, M. A., Rouabah, B., Kafi, M. R., and Toubakh, H. Health management using fault detection and fault tolerant control of multicellular converter applied in more electric aircraft system. Diagnostyka, 23, 2022. http://dx.doi.org/10.29354/diag/151039.
13. Nowakowski, T. and Komorski, P. Diagnostics of the drive shaft bearing based on vibrations in the high- frequency range as a part of the vehicle's self-diagnostic system. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 24(1):70–79, 2022. http://dx.doi.org/10.17531/ein.2022.1.9.
14. Ochella, S., Shafiee, M., and Dinmohammadi, F. Artificial intelligence in prognostics and health man- agement of engineering systems. Engineering Applications of Artificial Intelligence, 108:104552, 2022. http://dx.doi.org/10.1016/j.engappai.2021.104552.
15. Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., and Doll´ar, P. Designing network design spaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10428– 10436, 2020. http://dx.doi.org/10.1109/CVPR42600.2020.01044.
16. Rodriguez-Picon, L. A., Mendez-Gonzailez, L. C., Perez-Olguin, I. J., and Hernandez-Hernandez, J. I. A study of the inverse gaussian process with hazard rate functions-based drifts applied to degradation modelling. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 24(3):590–602, 2022.http://dx. doi.org/10.17531/ein.2022.3.20.
17. Ruan, Y., Yuan, L., He, Y., Li, Z., Yuan, W., and Lu, L. Prognostics and health management for piezoresistive pressure sensor based on improved gated recurrent unit networks. Measurement Science and Technology, 33(11):115112, 2022. http://dx.doi.org/10.1088/1361-6501/ac81a0.
18. Sawczuk, W., Merkisz-Guranowska, A., and Rilo Canas, A.-M. Assessment of disc brake vibration in rail vehicle operation on the basis of brake stand. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 23(2):221–230, 2021. http://dx.doi.org/10.17531/ein.2021.2.2.
19. Taghiyarrenani, Z. and Berenji, A. An analysis of vibrations and currents for broken rotor bar detection in three-phase induction motors. In PHM Society European Conference, volume 7, pages 43–48, 2022. http://dx.doi.org/10.36001/phme.2022.v7i1.3332.
20. Vaiciunas, G., Bureika, G., and Steisunas, S. Rail vehicle axle-box bearing damage detection considering the intensity of heating alteration. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 22(4): 724–729, 2020. http://dx.doi.org/10.17531/ein.2020.4.16.

21. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. Eca-net: Efficient channel attention for deep convolutional neural networks. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11534–11542, 2020. http://dx.doi.org/10.1109/CVPR42600.2020.01155.

22. Wen, X., Lu, G., Liu, J., and Yan, P. Graph modeling of singular values for early fault detection and diagnosis of rolling element bearings. Mechanical Systems and Signal Processing, 145:106956, 2020. http://dx.doi.org/10.1016/j.ymssp.2020.106956.

23. Xue, B., Xu, Z.-b., Huang, X., and Nie, P.-c. Data-driven prognostics method for turbofan engine degradation using hybrid deep neural network. Journal of Mechanical Science and Technology, 35(12):5371–5387, 2021. https://doi.org/10.1007/s12206-021-1109-8.

24. Yao, Z., He, D., Chen, Y., Liu, B., Miao, J., Deng, J., and Shan, S. Inspection of exterior substance on high-speed train bottombased on improved deep learning method. Measurement, 163:108013, 2020. https://doi.org/10.1016/j.measurement.2020.108013.

25. Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. Barlow twins: Self-supervised learning via redundancy reduction. In International Conference on Machine Learning, pages 12310–12320. PMLR, 2021.https://doi.org/10.48550/arXiv.2103.03230

26. Zhao, L., Zhu, Y., and Zhao, T. Deep learning-based remaining useful life prediction method with transformer module and random forest. Mathematics, 10(16):2921, 2022. http://dx.doi.org/10.3390/ math10162921.

27. Zhou, F., Xia, L., Dong, W., Sun, X., Yan, X., and Zhao, Q. Fault diagnosis of high-speed railway turnout based on support vector machine. In 2016 IEEE International Conference on Industrial Technology, pages 1539–1544. IEEE, 2016. http://dx.doi.org/10.3390/math10162921.